

# Ch 6.2: Shrinkage - The Lasso

Lecture 18 - CMSE 381

Prof. Elizabeth Munch

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Mon, Oct 15, 2024

# Announcements

## Last time:

- Ridge Regression

## This time:

- The Lasso

## Announcements:

- HW # 5 due Friday
- Exam next week!

Lec #	Date			Reading	HW
12	Mon	9/30	Leave one out CV	5.1.1, 5.1.2	
13	Wed	10/2	k-fold CV	5.1.3	
14	Fri	10/4	More k-fold CV,	5.1.4-5	
15	Mon	10/7	k-fold CV for classification	5.1.5	
16	Wed	10/9	Subset selection	6.1	HW #4 Due Weds 10/9
17	Fri	10/11	Shrinkage: Ridge	6.2.1	
18	Mon	10/14	Shrinkage: Lasso	6.2.2	
19	Wed	10/16	Dimension Reduction	6.3	
20	Fri	10/18	Overflow, Possibly more dimension reduction?		HW #5 Due Fri 10/18
	Mon	10/21	No class - Fall break		
	Wed	10/23	<b>Review</b>		
	Fri	10/25	<b>Midterm #2</b>		

# Section 1

Last time

# Goal

- Fit model using all  $p$  predictors
- Aim to constrain (regularize) coefficient estimates
- Shrink the coefficient estimates towards 0

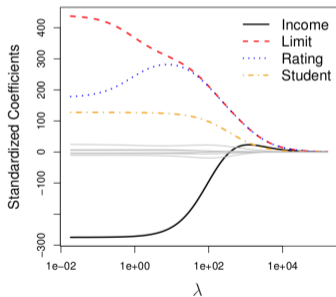
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

# Ridge regression

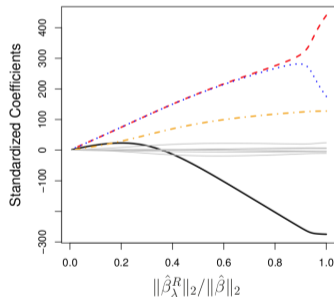
**Before:**

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2$$



**After:**



## Scale equivariance (or lack thereof)

**Scale equivariant:** Multiplying a variable by  $c$  ( $cX_i$ ) just returns a coefficient multiplied by  $1/c$  ( $1/c\beta_i$ )

**Solution: standardize predictors**

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$

- Least squares is scale equivariant
- Ridge regression is not

## Section 2

### The Lasso

## Same goal as before

- Fit model using all  $p$  predictors
- Aim to constrain (regularize) coefficient estimates
- Shrink the coefficient estimates towards 0

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$



# The lasso

**Least Squares:**

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

**Ridge:**

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2$$

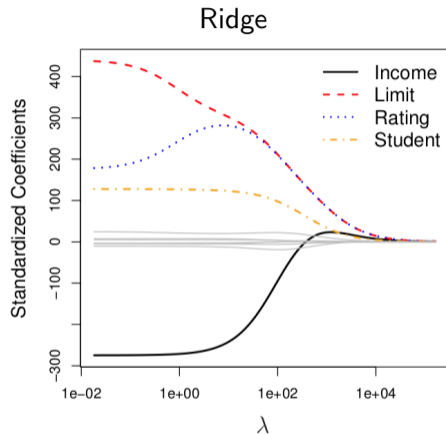
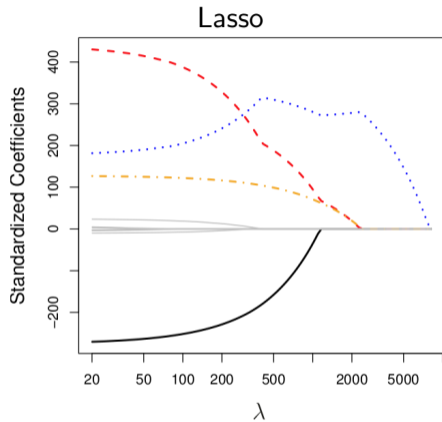
**The Lasso:**

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = RSS + \lambda \sum_{j=1}^p |\beta_j|$$

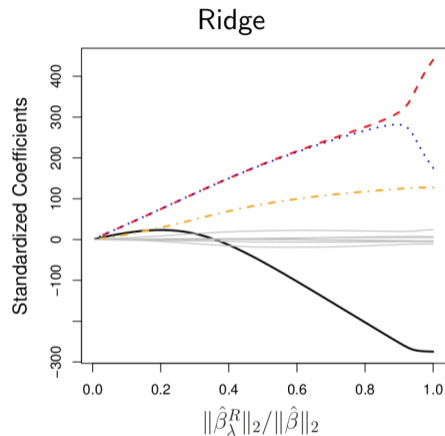
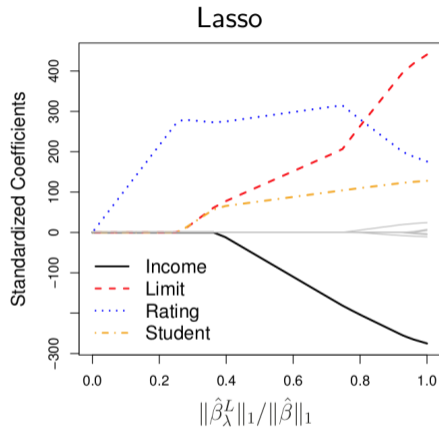
## Subsets with lasso

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$$

# An example on Credit data set



# More example on Credit data set



# Scale equivariance (or lack thereof)

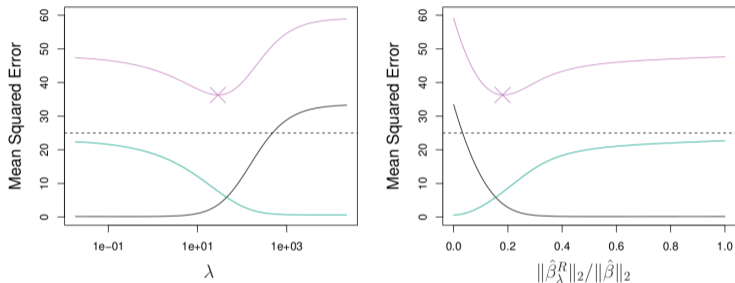
**Scale equivariant:** Multiplying a variable by  $c$  just returns a coefficient multiplied by  $1/c$

Least squares **is** scale equivariant.  
Ridge/Lasso **are very much not**.

**Solution: standardize predictors**

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$

# Bias-Variance tradeoff

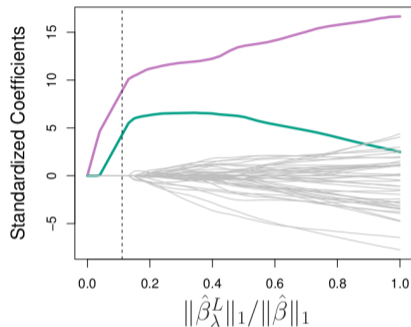
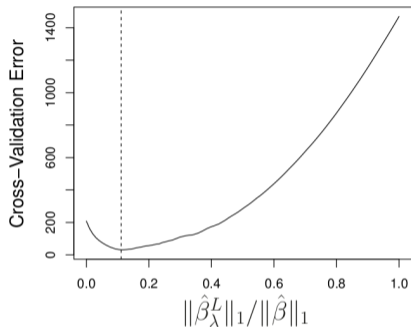


Squared bias (black), variance (green), and test mean squared error (purple) for simulated data.

## Using Cross-Validation to find $\lambda$

- Choose a grid of  $\lambda$  values
- Compute the ( $k$ -fold) cross-validation error for each value of  $\lambda$
- Select the tuning parameter value  $\lambda$  for which the CV error is smallest.
- The model is re-fit using all of the available observations and the selected value of the tuning parameter.

# 10-fold CV choice of $\lambda$ for lasso and simulated data





# Coding example

# Ridge vs Lasso

**Ridge Regression:**

**Lasso:**

## Least Squares:

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

## Ridge:

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2$$

## The Lasso:

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = RSS + \lambda \sum_{j=1}^p |\beta_j|$$

# Next time

Lec #	Date			Reading	HW
12	Mon	9/30	Leave one out CV	5.1.1, 5.1.2	
13	Wed	10/2	k-fold CV	5.1.3	
14	Fri	10/4	More k-fold CV,	5.1.4-5	
15	Mon	10/7	k-fold CV for classification	5.1.5	
16	Wed	10/9	Subset selection	6.1	HW #4 Due Weds 10/9
17	Fri	10/11	Shrinkage: Ridge	6.2.1	
18	Mon	10/14	Shrinkage: Lasso	6.2.2	
19	Wed	10/16	Dimension Reduction	6.3	
20	Fri	10/18	Overflow, Possibly more dimension reduction?		HW #5 Due Fri 10/18
	Mon	10/21	No class - Fall break		
	Wed	10/23	<b>Review</b>		
	Fri	10/25	<b>Midterm #2</b>		